Perception of Localized Features During Robotic Sensorimotor Development

Giagkos, Alexandros; Lewkowicz, Daniel; Shaw, Patricia; Kumar, Suresh; Lee, Mark; Shen, Qiang

Published in:
IEEE Transactions on Cognitive and Developmental Systems

DOI:
10.1109/TCDS.2017.2652129

Publication date:
2017

Citation for published version (APA):
Giagkos, A., Lewkowicz, D., Shaw, P., Kumar, S., Lee, M., & Shen, Q. (2017). Perception of Localized Features During Robotic Sensorimotor Development. *IEEE Transactions on Cognitive and Developmental Systems*, *9*(2), 127-140. https://doi.org/10.1109/TCDS.2017.2652129

Perception of Localized Features During Robotic Sensorimotor Development

Alexandros Giagkos, Daniel Lewkowicz, Patricia Shaw, Suresh Kumar, Mark Lee, and Qiang Shen

***Abstract*—The understanding of concepts related to objects are developed over a long period of time in infancy. This paper investigates how physical constraints and changes in visual perception impact on both sensorimotor development for gaze control, as well as the perception of features of interesting regions in the scene. Through a progressive series of developmental stages, simulating ten months of infant development, this paper examines feature perception toward recognition of localized regions in the environment. Results of two experiments, conducted using the iCub humanoid robot, indicate that by following the proposed approach a cognitive agent is capable of scaffolding sensorimotor experiences to allow gradual exploration of the surroundings and local region recognition, in terms of low-level feature similarities. In addition, this paper reports the emergence of vision-related phenomena that match human behaviors found in the developmental psychology literature.**

***Index Terms*—Biologically inspired feature extraction, conceptual learning through development, generation of representation during development, motor system and development, multimodal integration through development, robots with development and learning skills, visual system and development.**

I. Introduction

THE ABILITY to track and identify different visual objects is extremely useful for humans in order to perform adaptive behaviors in constantly changing environments. This ability relies on the capacity of a cognitive system to capture and maintain stable representations of separated entities from a continuous streams of visual input. One major challenge for cognitive robotics is the learning of perceptual skills and, in turn, the capacity of artificial systems to transform *percepts* into *concepts*. This paper considers how models of human development can be adapted and applied to robotics to start to address this challenge.

The human world consists of extended surfaces populated with distinct and separable objects. Gibson's ecological theory of perception [1] is based on the fact that the light, as received by the eyes, is not random and disorganized. Rather it reflects how the physical environment is structured. It bounces off objects in invariant ways, depending on their physical characteristics and situation in the layout. According to Gibson, humans perceive the object world primarily by detecting this invariant information. Human infants are perceptive to the object world, making some basic sense of it despite their marked visual immaturity during the first post-natal months [2].

The main objective of this paper is not to redefine the concepts of objects on a perceptual or psychological level. Instead, this contribution is an attempt to implement the necessary components for the processing of object qualities in embodied agents, whose cognitive processes are defined with simple rules and partially predetermined. As a result, this paper is asking the question about the understanding of the physical proprieties of stimulating regions in the scene through the systematic observation of their underlying structures. These apparent structures are a complex combination of perceptual scale changes during development and fortuitous repetitions in the visual input while observing a natural scene. Indeed, the opportunity to explicitly model some early properties of the human visual system in a developmental robotic platform enables the investigation of the effects of early visual and sensorimotor development on the understanding of objects. Moreover, patterns of sensorimotor contingencies in the representational structures of an agent's cognitive system during its interaction with the environment can be directly observed and described.

Applying the proposed architecture and following a longitudinal approach, an embodied agent is shown to be able to acquire visual experiences to scaffold its understanding and, in turn, build knowledge of its immediate world. This is achieved online and in a computationally inexpensive manner, with no prior knowledge or supervision. The results highlight that when maturing according to the proposed developmentally plausible time-line, an agent is able not only to gain sensorimotor experiences by visually interacting with the environment, but it is also able to facilitate recognition of stimulating regions as well as detecting dramatic changes that may occur. While currently limited to visual stimuli, this approach lays the foundation to expand to multimodal representations that can gradually be developed and refined through the accumulation of experiences, the range of which is steadily increased over time.

This paper is organized as follows. Section II describes the time-line of development observed in infants for both

Manuscript received March 31, 2016; revised September 6, 2016 and November 7, 2016; accepted December 16, 2016. Date of publication February 6, 2017; date of current version June 8, 2017. This work was supported by the U.K. Engineering and Physical Sciences Research Council under Grant EP/M013510/1 (Developmental Algorithms for Robotics; Understanding the World of Objects, Interactions and Tools). *(Corresponding author: Patricia Shaw.)*

The authors are with the Department of Computer Science, Aberystwyth University, Aberystwyth SY23 3AU, U.K. (e-mail: phs@aber.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCDS.2017.2652129

sensorimotor and feature perception, followed by Section III which discusses the development's application to a robotic platform. Section IV presents experimental results evaluating the developmental progression and localized region recognition, finishing with a discussion in Section V.

II. Background

A. Time-Line of Visual Development

In most mammal species, visual development is delayed due to the limited visual input in the womb as well as the limited availability of patterns being perceived. One major discovery of the early visual neuroscience is the example of visually deprived cats with one eyelid sutured shut at birth or during development and the discovery of sensitive periods for the maturation of the primary visual cortex [3]. In humans, the experiments of Maurer *et al.* [4] demonstrate that acuity does not improve post-natally until the nervous system receives patterned visual input. For deprivation (surgical removal of cataractous lens, leaving the eye with no means to focus on images) up to nine months after birth, acuity remains close to the newborn's level. However, at this age, patterned visual input can alter the nervous system rapidly and sufficiently to support better acuity as early as 1 hour later and to induce further improvement over the next months [4].

Newborns' visual acuity or their ability to detect variations in fine detail, is approximately 40 times poorer compared to adults with healthy vision. It is only by the end of the first year that infant visual acuity approaches the adult level [5]. No matter what method is used to measure it, visual acuity is poor at birth; the smallest stripes to which newborns respond are approximately 40 times larger than what can be resolved by individuals with adult vision [6]. There is at least a fivefold improvement in acuity by six months of age, although it takes several more years for acuity to reach adult levels [7].

Poor acuity at birth is likely caused both by immaturities in the size and arrangement of retinal cones and by additional limitations beyond the retina [8]. The rapid improvement in the first six months reflects, in part, the development of foveal cones so that they filter out less information and allow finer and finer detail through to tune cells in the visual cortex [8], [9]. Maurer and Lewis [6], Mayer *et al.* [10], and Courage and Adams [7] provided an accurate time-line to simulate the important acuity changes in the first six months of life.

The visual field is restricted at birth in comparison to the visual field of adults and shows considerable expansion over the first months of age. Later, the visual field expands from approximately 75% of adults at 7–9 months to nearly 100% by 2.5 to 3 years [11]. Despite the few measurements and variations in methods (e.g., intensity of lights and size of stimuli), it has been generally found that babies orient toward targets from between 0° and 30° at birth to between 90° and 105° at six months [12]. A detailed review of these results in the first month of age is provided in [13]. For this paper, inspiration is take from these time-lines to provide an accurate simulation of infant visual input for the system (see Table I, as it will be explained in Section III-C).

Infant color vision is poor at birth and most psychophysical experiments agree that infant color vision emerges between the age of three weeks and three months [14]. The overall insensitivity of infants to contrast is likely to provide an explanation of the poor color vision of infants [15].

Before two months of age, infants do not consistently demonstrate the ability to discriminate stimuli that differ in hue only. Older infants, however, can reliably discriminate these differences in color [16], [17]. However, the extent to which younger infants use pattern and color features when they reason about physical events is not immediately apparent. Although there is evidence that surface features can, in some instances, influence infants' performance on physical reasoning tasks [18], their use by infants has not been systematically explored. The lack of evidence may suggest that, in many physical situations, surface features are simply irrelevant to the outcome of the event. For example, the color of the ball would not alter its ability to fit in the opening [19]; likewise, no matter if the rabbit's body was striped, dotted, or plain, it would not affect its ability to appear in the window [20]. Hence, infants may not be practised at determining when surface features provide important information and at using this information effectively [21].

According to [2], the way in which poor vision of neonates might help constrain them to infer and represent some invariant principles and to figure out the physical world, is still an open and stimulating question. "Like astrophysicists theorizing about invisible worlds by inferring from the poor visual information provided by telescopes, infants would likewise be constrained to infer from the poor perceptual information they are able to gather." Whereas this question cannot be directly tested in human infants, it is now possible to perform experiments on robotic surrogates. For example, a system that explicitly modeled the formation of object percepts through distinctive stages of visual development, might be able to shed some light on the invariant underlying structure of the visual input despite the massive developmental changes briefly exposed above.

However, one important question regarding such a system is the amount of *a priori* information and its initial abilities as a starting point for the vision development to be able to systematically investigate its effect on object perception. In order to avoid the nativist/empiricist debate which is out of the scope of this paper, focus is put on the most common abilities that are mentioned in the developmental psychology literature and some of the best arguments in favor of a few prenatal visual abilities even prior to the onset of the patterned visual input [22].

First, the ocular-motor system is sufficiently functional to react to high saliency targets. The classical types of particularly salient targets are high-contrasts edges, motion, and face (not relevant to this contribution). Thus, in the presented model, camera input is separated into four channels: 1) brightness; 2) color; 3) motion; and 4) edges.

Second, the organized activity in visual pathways from early on, that contribute to retinotopic "mapping" preserve the sensory structure, e.g., relative positions of neighboring points of visual space from retina through the thalamus,

visual cortex and higher visual regions. Thus, correlated inputs would remain coupled and dissimilar inputs could become dissociated. This has two consequences for the proposed model:

1) an ability to link the visual features close to each other in regard to their relative location in the developing notion of the immediate physical space;
2) the idea that "edges" should be defined as simple, local interactions in the input enabled by a very simple feed-forward calculation.

Third, there is evidence of a very simple but still present system of short term memory at birth. For example the effects of habituation, recognition and differentiation of familiar or novel patterns and the interest toward new stimuli is often exploited by experimentalist to indirectly measure other cognitive abilities by comparing how long a baby will fixate a target. In the presented model, this can be seen as both ability to collect simultaneous information about targets and to confirm its persistence if this information is already seen.

A newborn baby is equipped with perceptual and cognitive mechanisms sufficient to begin the process of learning about objects by detecting edges, tracking motion, recognizing familiar items, discriminating items presented simultaneously or in sequence, and so forth. Having these concepts in mind, the initial stage and the incremental changes of the proposed model are presented in the next sections.

III. Sensorimotor Control and Region Perception Through Staged Development

In this section, a system able to build sensorimotor as well as region perception experiences according to the human developmental time-line discussed above, is presented. Reflecting the system's modular design, each component is described separately.

A. Robotic Vision

A vision module is developed to model a close approximation of the infant's visual ability at each month of development, in connection to the psychological literature. It is designed to provide two core functionalities.

1) A way to configure filters that alter the field of view (FOV), acuity and contrast of input images according to the developmental time-line.
2) Four low-level feature extraction mechanisms that detect features and measure their quantities.

Currently, the module is capable of locating stimulating targets based on their color, level of brightness, motion activity as well as their edges. As long as they satisfy the module's minimum region thresholds, red and green targets are identified in each image. Color detection is achieved by comparing the hue, saturation, and value (HSV) attributes of each pixel against the range that define each color in the HSV color space. Subsequently, the centroid of each target as well as its size (in pixels) is reported, followed by the mean hue and saturation values. This approach not only allows the system to recognize colored targets per se, but it also offers the ability to ultimately distinguish between targets of the same color, based on the detailed color information.

In the same vein, brightness detection is achieved by measuring the average value attribute of each of the identified targets and matching it with an acceptable range in the HSV color space. In order to reduce environmental noise, regions smaller or larger than predefined size thresholds are ignored, making sure that only reasonable targets are seen within a very noisy environment. The brightness filter reports the centroids and sizes as well as the average value of each target.

For edge detection, the module employs the Canny edge detection algorithm [23]. Noise is reduced by excluding regions that do not satisfy size thresholds and are not clearly defined by their detected edges, that is, not all of the pixel points of the region belong to the same convex set.[1] Additionally, once such a target region is identified, the vision module calculates its perimeter and the Euclidean distance between pairs of extreme points, namely top and bottom as well as left and right extremities. The horizontal and vertical distances as well as the perimeter of the region, clearly identified by its edges, are reported. In the proposed system, this edge detection is considered as a precursor for a more mature, shape-detection ability, as it gives some insight related to the shape of a potential object in the scene.

Finally, the motion extraction mechanism compares two consecutive images. To decrease resolution, both images are blurred and turned into gray scale. The comparison identifies the centroids as well as the sizes (in pixels) of the regions that differ. This approach is suitable for seamlessly detecting any target region that is perceived due to some motion activity, including flashing lights.

Notice that the vision module provides only low-level image processing and is capable of reporting feature-related information for multiple targets continuously. The rest of the system is responsible for analyzing, storing and associating the data. It is through the representation of the data and the correlations being made while the cameras observe the environment that the system builds its understanding about interesting areas in the scene.

The data structures are described in the next section.

B. Maps, Fields, and Links

Both eye and head control learning as well as the region perception are based on a mapping module. Several multidimensional structures or maps are used in order to represent sensorimotor as well as feature-related search spaces within the system. Instead of points, each map consists of small overlapping regions or fields, whose center points and radii allow them to represent closely related values within the particular search space.

Although within the developing brain, the structure of the receptive fields and neurons is closely related to the acuity and FOV, within this system fields represent regions of equivalence that could be loosely considered as receptive fields. Notice that this paper does not focus on producing a neurally

[1]This is a connected set in which lines can be drawn between any two points without leaving the set.

accurate model, rather it takes inspiration from psychology and neuroscience.

In the maps technology, fields can be connected using explicit links. More specifically, when a stimulating target is identified by the vision module (described in Section III-A), a representation of it is created in the retina map. This map arranges fields in a polar grid, where each radius is proportional to the distance between the centroid of the corresponding field and the center of the image, based on images of resolution 360 × 240. The motor values of the eyes and the head are also accommodated in two polar motor maps, whose dimensions are defined by their joints' range of values (i.e., pan and tilt joints).

Through the process of motor babbling, links between retina and eye motor as well as retina to head motor maps are gradually learned [24]. Links are used not only to make correlations between regions of different spaces, but also to allow traversing between experienced regions within a single map. In the context of the sensorimotor control, links connect retina fields with their correlated eye motor fields, associating relative eye motor movements for fixation with what is visually observed. Links are also used to construct eye and head movement trajectories utilizing relative movement information found in the motor fields, toward successful fixations on targets. In more detail, when a stimulus appears in the retina, the corresponding retina field is activated. By accessing it, links to the corresponding eye and head motor fields point to the appropriate adjustments to apply to the current eye and head configuration in order to perform a saccade toward the target. If no previous information exists, neighboring retina fields are used instead, if available. That ensures that the target is brought closer to the foveal area. However, if due to insufficient number of fields the target cannot be centered, the learning algorithm triggers a random babbling behavior to promote learning of new sensorimotor experiences.

During head motor movements the vestibulo-ocular reflex acts to maintain the eye fixation on the target. The internal mechanisms and the algorithms of this learning approach as well as the formulas for calculating the gaze shift (i.e., contribution of eyes and head to perform successful fixations) are found in [24]. Notice that in the eye and head motor maps, the number of new links decreases over time as existing ones are used, giving a level of body-related saturation in the mapping module [25].

Similar to the learning of sensorimotor control, localized region perception uses map structures in order to store and correlate collected features from targets that define aspects of a small region within the observable space, for example, the immediate environment of the robotic platform containing an object with multiple visual features detectable from it. Four Cartesian feature maps exist in the system.

1) A 2-D map is used to accommodate feature values related to the *brightness* of located stimulating targets. The dimensions of the map are defined by the range of the value element of a pixel in the HSV model (0–255) and the size of the target region in pixels (0–5000). Feature fields in the brightness map represent targets that differ in size and in level of brightness. Their radius is fixed to 5 units, due to the narrow range of values expected for the value element. That allows the system to better distinguish between targets of different brightness, yet of similar size.
2) A *color* map is designed to represent color information for colored targets. Its fields are defined by the hue and saturation elements, with ranges 0–180[2] and 0–255 defining the map dimensions, respectively. Similar to brightness, the radius of each field is fixed to 5 units.
3) A single dimension *motion* map is able to accommodate information about mobile targets. Its dimension ranges from 0–5000 and corresponds to the size of a target region in number of pixels. Here, due to the large range of dimension and the expected fluctuation related to region sizes, the radius of each field is 20 units.
4) An *edges* map is designed to represent information related to target regions defined by their edges. This is a 3-D map where fields are defined by the perimeter of the identified target region as well as its horizontal and vertical distances, as they are received from the vision extraction mechanism. The field radius used in this map is 20 units.

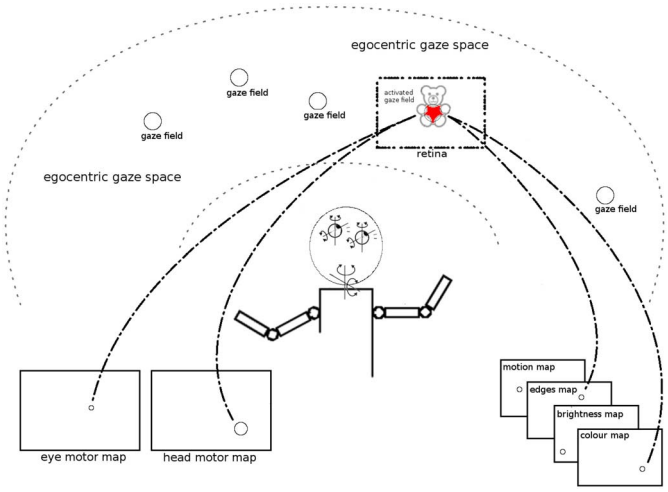

Fig. 1. Illustration of the gaze space and associations between sensorimotor control maps as well as feature maps linked to gaze space for localized region perception. The agent fixates on a stimulating area (red patch) and extracts any available low-level information (color and edge features).

Along with the feature maps, there exists a Cartesian map to represent an ego-centric gaze space consisting of gaze fields, where spatial locations of targets relative to the robotic platform are recorded. This map is 2-D and its fields are defined by the gaze pan and tilt coordinates as described by the gaze orientation of the eyes and the head. Through the process of feature collection from targets identified by vision, feature fields are created and linked to the particular gaze fields, which in turn are excited to represent and locate stimulating regions that are both in and out of the retina. The system is able not only to know the spatial location of stimuli previously experienced, but also to describe each stimulating region by

[2]Notice that the typical hue range is scaled down, due to the use of OpenCV for image processing.

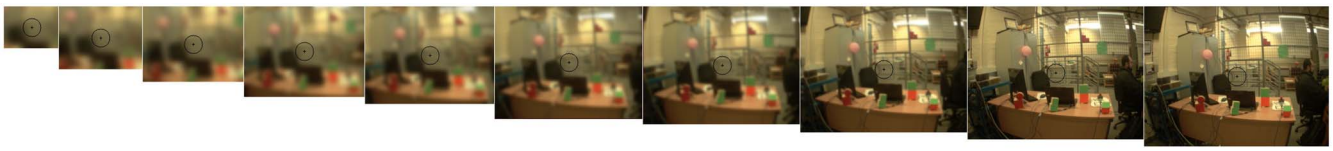

Fig. 2. Effect of the staged vision development on perceiving the world. The order of the development is depicted from months 1 (left) to 10 (right). The black circle in the center of each image marks the fovea of the eye, with the cross in the center being the fixation point.

TABLE I
STAGED VISION DEVELOPMENT

Month	1	2	3	4	5	6	7	8	9	10
FOV	.30	.45	.55	.65	.70	.80	.85	.90	.95	1.0
Acuity	40	35	30	25	20	15	10	5	1	1

TABLE II
SATURATION TYPES AND THRESHOLDS

	Eyes		Head		Combined
Month	All	Re-used	All	Re-used	
1	40	10	–	–	–
2	80	20	–	–	–
3	160	40	–	–	–
4	–	–	80	30	–
5	–	–	100	50	–
6	–	–	140	70	–
7	–	–	180	90	1
8	–	–	–	100	2
9	–	–	–	120	2
10	–	–	–	130	3

revisiting the collection of features it has stored. The system is better illustrated in Fig. 1 where the agent is shown to fixate on a particular stimulating region in its immediate environment, activating the corresponding gaze field. Notice that the retina of the agent follows the agent's gaze within its egocentric space. The head configuration (i.e., eye and head motor fields) is associated with the activated gaze field. In terms of perception, the agent is shown to extract color and information for edges, activate the appropriate feature fields and link them with the gaze field of the fixation.

C. Vision Development

Vision development is designed to progress in stages (reflecting months) according to Table I, with both FOV and acuity affecting the visual perception of the system. Note that the maximum FOV of the eye cameras used in this paper approximately correlates to that of an one year old child, rather than the full adult level described in Section II-A. The level of acuity was measured using a series of infant acuity tests, based on the examples provided by the Vision Research Group at Ulster University [26].

Acuity is simulated by applying a smoothing operation to the input image. The operation involves linear convolution with a Gaussian kernel, the size of which is determined by the aperture width, represented in the model by the acuity parameter. The range of the latter is 1–100, with 100 being dramatically affected and 1 resulting in applying no smoothing. Again, the values of the acuity parameter in Table I are carefully chosen to match the time-line discussed in Section II-A. Fig. 2 depicts the effect of the vision development between months and the impact it has on the quality and level of detail, thus the perception ability of the environment through the system cameras.

D. Sensorimotor Control Development

As the impact of the vision development to both eye and head sensorimotor control and learning of feature clusters is investigated in this paper, it is important to define the conditions and constraints that make the system experience the infants' development at each month. For this purpose a lift constraint, act, and saturate [27] approach is followed. Apart from the vision staged development described in Section III-C, a head movement constraint is applied and lifted at just after the third month of age. This is to reflect the infants physical inability to effectively control the neck muscles till later months.

In order to define a level of saturation for each month, four performance metrics are considered:

1) the number of links that exist between retina and eye motor maps;
2) the number of links between retina and head motor maps;
3) the number of those links that are reused;
4) the number of saccades that are performed using links that involve both eye and head maps simultaneously.

These metrics are used as indicators for the maturity of the system, because they are attributed to the particular visual as well as physical conditions of the system at each month.

A large number of links between the retina and the eye motor maps suggests an equally large number of eye-related sensorimotor experiences, collected as a result of frequent eye babbling and successful fixations. As the number of successful eye saccades increases, the development of the sensorimotor control also increases, rendering the system mature enough to progress to the next month.

Once links are added, repeated use of these links acts to confirm their correctness and thus proving their success in leading to fixations between targets. As not all of the links are ultimately reused, the number of those that are is significantly less than the total number. As a result, the sensorimotor control performs less motor babbling and more targeted eye movements as more links are reused. This is due to the utilization of successful links, allowing the eyes' position to change according to consecutively linked fields in the motor space.

"Combined" saccades are the saccades where links from both eye and head control are utilized in order to change between two successfully fixating gaze directions. Combined saccades are themselves an indication that the maturity level

Algorithm 1 Regional Perception Mechanism

Require: A successful saccade leading to target T
if *isFixating*(T) **not** *true* **then**
  return // lost stimulus
end if
$N \leftarrow$ *getNeighbouringTargets*(T)
for $\forall n \in N$ **do**
  $\mathbf{t} \leftarrow$ *determineType*(n)
  access $f_t \in F_t$ and $g \in G$
  $l \leftarrow$ *createLink*(g, f_t)
  $l.confirm \leftarrow l.confirm + 1$
  $l.timestamp \leftarrow timeNow$
end for
$g.confirm \leftarrow g.confirm + 1$
$g.timestamp \leftarrow timeNow$

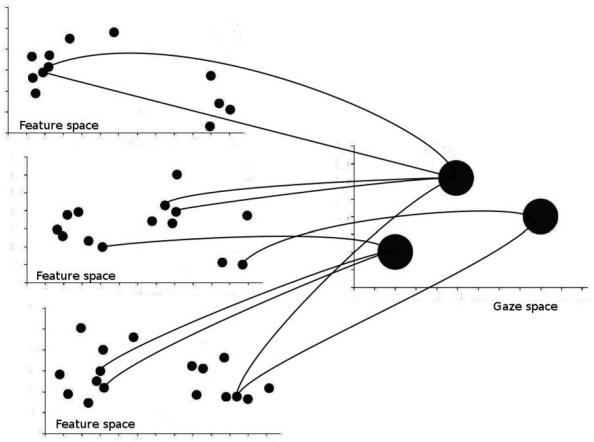

Fig. 3. Illustration of mappings between feature fields (e.g., color, brightness, etc.) and gaze fields. These are created as the system explores, extracting clusters of localized visual features in regions of gaze space.

of the system is such, that enough reused links exist for both eye and head control, demonstrating a level of gaze control.

Table II summarizes the saturation types and thresholds that are used for each month of development. Notice that after month three, the development of the eye motor control is close to a satisfactory level of saturation. In fact, half of the motor map is populated and the number of reused links is enough to allow direct transitions of the eyes between two fixations. Between months 4 and 7, saturation is mostly concentrated on the head motor control which, taking advantage of the progress the eye motor control has made thus far, is expected to develop quickly. Given the particular visual constraints at each month, the thresholds are finalized based on the inability of the system to achieve better performance.

E. Modeling Regional Perception and Recognition

It is clear that both learning of sensorimotor control and regional perception of feature clusters depend on the effective population of the underlying maps, as well as the number of links that connect corresponding fields together. The two learning mechanisms are designed to work in parallel, so that the development of the one directly affects the development of the other. As previously shown in [24], [25], and [28], motor babbling has been effectively used to drive the discovery of new experiences between a robotic platform's sensors and motor. This mechanism has been previously evaluated and shown its effectiveness in being a competent approach to achieve robotic sensorimotor learning. In this paper, focus is given on the process of: 1) discovering such sensorimotor experiences during the staged development of infant's vision and 2) building knowledge related to stimulated regions within the gaze space of the system.

In particular, gradually building knowledge about stimulated regions primarily consists of feature extraction, analysis, manipulation and association of feature data. Given that the vision is stimulated by targets consisting of features of certain types (i.e., brightness, color, motion, and edges), when the eyes fixate on a region, e.g., as a result of a successful saccade performed by the sensorimotor control, the system is expected to extract at least one feature within the region of fixation. The latter is considered as the eye's fovea and has a fixed radius of 10% of the image width.

The system is capable of collecting neighboring features related to the same small region in the gaze space. In spite of the single feature extraction, which confirms the existence of a stimulus and in turn excites the corresponding gaze field, the region perception mechanism is designed to extract neighboring features in an attempt to learn as much as possible during a single fixation. Neighboring features are those within the foveal region. That is, if two or more feature-related targets are found within the fovea, they are considered to be neighbors. The regional exploration mechanism is shown in Algorithm 1. The results of the iterative process that aims at gathering and analyzing extra stimulus found within the fovea with $\mathbf{g}$ is depicted in Fig. 3, where excited fields from feature spaces (e.g, color, brightness, etc.) are associated with gaze fields in the gaze space.

When the eyes fixate on a stimulating region, as a result of an identified target $\mathbf{T}$, information is collected. For example, if a red target is responsible for eye fixation, the color information is used to excite an existing color field $\mathbf{f}$ in the feature map $\mathbf{F_t}$ ($\mathbf{f} \in \mathbf{F_t}$), where type $\mathbf{t}$ is now color. Notice that in order to facilitate the identification of the fixated target in the foveal area, the feature type is mainly used for target matching. Furthermore, if no existing color field is present, the mapping system will create a new one in order to accommodate the color information. To associate the feature information with the current position in the gaze space, a link $\mathbf{l}$ is made to connect $\mathbf{f_t}$, that is a feature field of type $\mathbf{t}$ (e.g., $\mathbf{t}$ is "color" for fields in the color map) with the appropriate gaze field $\mathbf{g} \in \mathbf{G}$, with the latter being the egocentric gaze space. If such a link is already present, then it is reused to enhance the particular visual experience in $\mathbf{G}$.

As previously stated, the sensorimotor control and regional perception mechanisms depend on each other, with the first acting as the driving force of the agent's attention. While learning sensorimotor control, the agent employs random movements as well as calculated saccades resulting from previously learned links. Both drive the attention of the eyes to stimulus that, when in the center of the fovea, allows feature extraction to be performed. Notice that at this current stage of

the architecture, there is no excitation mechanism to drive the attention of the system to particular stimuli, rather learning of the environment is dictated purely by the sensorimotor control mechanism.

An exception to the above is when the agent is forced to drive its attention to similar regions. As this paper focuses on region perception, designing a system capable of perceiving similar regions within the immediate space is taken into consideration. Driven by a simple mechanism that identifies similarities in regions, the system allows the artificial infant to shift its gaze toward those regions that are found similar.

The concept of similarity as well as the mechanism to compare two regions depend on the low-level feature collection. Remember that feature maps are employed to accommodate values received from the four vision inputs when the eyes fixate on a region of interest. For example, when a bright object is placed in front of the infant it is perceived as an interesting region that contains (at least) a brightness related target. When the eyes fixate on it, its value and size create or trigger existing fields in the brightness map which are linked with a gaze field that represents the particular region's position in the gaze space. Notice that several brightness fields may be linked to a particular region, as a result of being seen and perceived multiple times. In addition, similar features may have been seen in different positions resulting in links to multiple gaze fields.

Therefore, comparing two regions in terms of a low-level feature is achieved by: 1) calculating the overall feature information that is associated with each of the regions and 2) measuring the distance between the two. Following the mapping approach, the Euclidean distance between two average brightness fields in a brightness map is used to quantify how different two regions are with respect to brightness. Two overlapping average brightness fields is an indication that the two regions are similar in terms of brightness (i.e., value and region size).

Notice that the overall feature information is measured by calculating the weighted average of the feature fields found linked with a region, where the number of link confirmations are used as the weights for averaging. For instance, if the link between a gaze field and a particular brightness field is confirmed multiple times, it should have a stronger presence when calculating that region's brightness information

$$\bar{F}_t = \frac{\sum_{i=0}^{n}\left(F_{t(0)} \times \omega_i\right)}{\sum_{i=0}^{n} \omega_i}, \ldots, \frac{\sum_{i=0}^{n}\left(F_{t(k)} \times \omega_i\right)}{\sum_{i=0}^{n} \omega_i}. \quad (1)$$

Equation (1) is used for calculating the weighted average. Here, $\mathbf{t}$ is the type of the feature, $\mathbf{n}$ is the total number of feature links associated with $\mathbf{g}$, and $\mathbf{k}$ is the number of elements included in the particular feature field $\mathbf{F_t}$ according to the information it accommodates (e.g., $\mathbf{k} = 2$ for brightness; value and size elements). Hence, in the example of brightness information, $\mathbf{F_{t(0)}}$ would be equal to the value element of the first brightness field.

Looking for similar regions in the gaze space is shown in Algorithm 2, which is supplementary to Algorithm 1 to allow

Algorithm 2 Exploring Similar Regions

14: $S \leftarrow getSimilarRegions(g)$ // algorithm 3
15: **for** $\forall g' \in S$ **do**
16: $\quad applyHeadConfiguration(g')$
17: $\quad T' \leftarrow getClosestTarget(g')$
18: $\quad$ *repeat steps* $4-11$ *for* T' *and* g'
19: $\quad memoryDecay(g')$
20: $applyHeadConfiguration(g)$
21: $memoryDecay(g)$

Algorithm 3 Recognizing Similar Regions

1: $S \leftarrow 0$
2: $\bar{F_c}, \bar{F_b}, \bar{F_m}, \bar{F_e} \leftarrow avgFeature(g, \{c, b, m, e\})$
3: **for** $\forall g' \in G$ **do**
4: $\quad$ **for** $\forall t \in \{c, b, m, e\}$ **do**
5: $\quad\quad \bar{F_t}' \leftarrow avgFeatureField(g', t)$
6: $\quad\quad d_t \leftarrow dist(\bar{F_t}, \bar{F_t}')$
7: $\quad\quad$ **if** $d_t \leq F_t.radius$ **then**
8: $\quad\quad\quad S \leftarrow S \cup g'$
9: return S

Algorithm 4 Memory Decay

1: $L \leftarrow linksAssosiatedWith(g)$
2: **for** $\forall l \in L$ **do**
3: $\quad ts\_offset \leftarrow (timeNow - l.timestamp)$
4: $\quad$ **if** $(ts_offset > (Thresh \times l.confirm))$ **then**
5: $\quad\quad l.confirm \leftarrow l.confirm - 1$

region perception and recognition. Once all possible features within the foveal area are extracted, the agent is forced to recall similar regions that have been previously experienced, and drive its attention toward them, updating features recorded for each of them.

Algorithm 3 is utilized to return the regions whose feature information overlaps with the region associated with **g**. While iterating though similar regions, the system collects information about them, repeating the exploration mechanism. This is found to be very important as it allows the system to revisit previously discovered regions and identify any differences in the way they are perceived.

Notice that the perception of the environment changes according to the vision development, thus past and recent experiences are fixed when calculating the average feature information. If during early stages a region is visited multiple times, then the impact of the links made to the weighted average will be stronger. Through development, this impact is transferred to later stages rendering the system prone to making mistakes when comparing regions that are changed or perceived differently. Undoubtedly, the ability to forget in order to effectively use past experiences is important. This is achieved by enabling a memory decay mechanism at lines 19 and 21 in Algorithm 2. The internals of this mechanism are depicted in Algorithm 4, where the time when each link was last confirmed is examined. A time offset is calculated based on each link's confirmation number multiplied by an arbitrary fixed threshold (3 min). In this fashion, memories of features that are observed more will last longer and will have a higher contribution to the averaging. This simple technique allows

the system to forget feature correlations that are not persistent during the development or experienced only once.

The current design offers a level of data transparency that renders the system able to reason about specific regions in terms of their associated features. Given sufficient time and exposure to an environment with enough stimuli, data patterns start to emerge highlighting feature aspects related to the surface and shape of experienced regions. As more refined data is collected, the stronger the system's ability to describe what is visually observed becomes.

IV. Experiments and Results

This section is organized as follows. First, the experimental methodology is described, including the description of the environmental conditions and set up as well as the steps taken to conduct a longitudinal study using a robotic platform. Next, results related to the learning of sensorimotor control and the effect vision development has to its progress are reported. Finally, observations and results about the ability of the system to perceive the stimulating regions within the environment are given and analyzed.

A. Experimental Methodology

In order to evaluate the system, two experiments following a longitudinal approach using the iCub humanoid robot are conducted. The robot's vision is achieved by two DragonFly2 cameras of low resolution (320×240, 25 frames/s). The robot is placed in a laboratory with typical indoor light conditions. Several targets are placed in front of the robot in order to stimulate the vision and exercise the feature extraction mechanisms. Namely, red and green targets that vary in HSV values, shape and size are placed within the scene. Flashing lights, including the robot's own LEDs situated on its arms, are used to represent targets in motion. Finally, bright targets are naturally present due to high value element (considering the HSV model) of regions in the scene (e.g., pink and light green regions) and the illumination in the room. Thus, the robot is exposed to a high level of noise arising from a natural, highly dynamic environment.

The first experiment is designed as follows. For each stage of development, the learning of sensorimotor control and feature perception are performed in parallel. The duration of each stage is dominated by the corresponding saturation types and level thresholds found in Table II. As the development progressively moves from one stage to the other, data collected during previous stages is utilized, allowing the robot to gradually and cumulatively build on top of previous sensorimotor experiences. However, in terms of regional perception the data previously collected is not utilized. Rather, the system has to relearn its environment by observing the extracted features related to specific regions within its gaze space. As the focus of this experiment is: 1) to examine how the development of vision affects the sensorimotor learning and 2) to evaluate how the system progressively perceives the environment and scaffolds its knowledge about it during stages, no memory decay and comparison of similar regions are employed. Rather, the robot is exploring its surroundings purely based on its ability to extract low-level features from interesting regions, thus utilizing only Algorithm 1. Notice that no regions change between stages, as a result of moving objects around, and the motion filter is mainly triggered by the mobility effects of the robot, flashing lights and general noise in the environment.

It is also worth mentioning at this point that targets, independent of their type, are detected within most of the scene. Inevitably, the amount and type of information to be returned purely depends on the visual ability of the system at each stage. Having regions identified by different feature types being located very close to each other, is an expected phenomenon. For instance, a cubic object of both red and green facets within a region is characterized by at least two pieces of color plus edge information. Color-wise, information regarding stimuli of green and red regions will be linked to the same gaze field and as far as the system's ability to reason based on color is concerned, distinguishing between colors is subject to the number of link confirmations as well as the perceived level of color details.

Furthermore, none of the targets are placed within reachable distance. Depth, although factored out when defining the gaze space of the robot at this level (i.e., the gaze space is represented as a 2-D space), is responsible for the grouping of neighboring feature targets together, as they are misleadingly perceived by the individual cameras. When this situation occurs, it is considered as noise that the experimental configuration tolerates in order to simulate a realistic as possible environment with lack of depth perception in the first months.

At the end of each stage, data in the form of maps and links between their fields is stored and analyzed. Results in terms of the effect and the impact of vision development to learning of sensorimotor control and object perception are presented and discussed in the next section.

Finally, the results of a second experiment designed to evaluate the capacity of the system in recognizing similar regions and identifying changes related to low-level feature collection are included. The second experiment is conducted following a methodology similar to the first experiment; however, now the system's ability to perform simple memory management and to compare previous knowledge with newly collected features is evaluated. Hence, during experiment two, the system identifies similar regions within the scene as described in Algorithms 2–4. Results on the recognition of similar regions are discussed in Section IV-D.

B. Results on Sensorimotor Control

During the development of the vision, with FOV and acuity values moving closer to those of an adult, noticeable effects on the learning of sensorimotor control are observed. While saccading between two fixation points, the robotic eyes and head apply a series of motor configurations that are associated, or linked, with corresponding excited retina fields. The number of links to be utilized is proportionate to the eye and motor experiences the robot has previously acquired. Undoubtedly, the amount of experiences depends the system's ability to ultimately fixate, which in turn is firmly related to the accuracy and efficiency of the vision.

TABLE III
AVERAGE NUMBER OF EYE MOTOR LINKS THAT ARE REQUIRED TO COMPLETE A SACCADE

Month	1	2	3	4	5	6	7	8	9	10
Avg	4.2	3.6	3.4	2.4	2.3	1.7	1.5	1.4	1.26	1.1

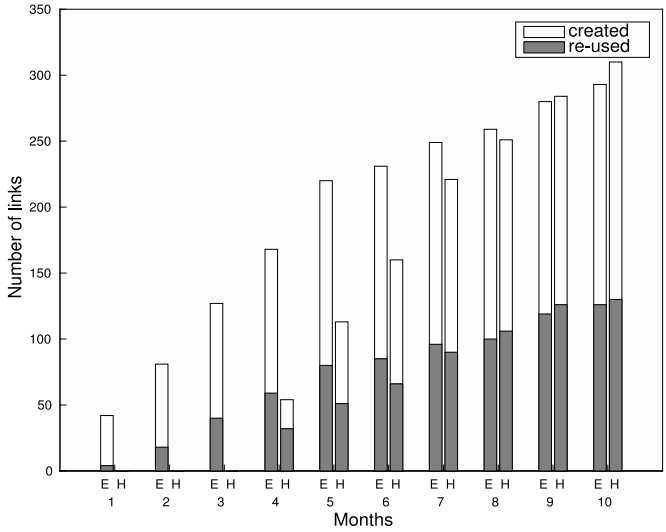

Fig. 4. Sensorimotor links between retina and eye/head motor spaces created at the end of each month. Filled regions depict the amount of reused links for eye (E) and head (H) control, respectively.

In this context, Table III shows the average number of steps, i.e., retina to eye motor links, that are required to complete a successful saccade. During the first months, the eyes need to make more steps in order to detect and fixate on a particular target, as a result of the underdeveloped vision. As expected, while the vision develops the number of steps decreases. Reused links between retina and eye motor maps are utilized to drive the robot's gaze so that the target is centered. The narrowed and blurry vision during the first month (as depicted at the far left of image 2), coupled with the locked neck restrict target detection, with an average of four links being required for each saccade. In the contrary, during month ten the robot is capable of performing single step saccades most of the time, an indication that a large number of links are developed even for remote targets.

In Fig. 4, the progress made during the learning of the sensorimotor control is analyzed in terms of the number of reused links, as compared to their total number of links created. Results for both eye and head maps are depicted. Notice that no head data is present until month three, when the neck lock constraint is lifted, and that the learning follows a cumulative fashion. The ratio between the reused and total links changes according to the maturity level, with the system being able to reuse approximately 40% of the total eye and head links that it progressively generates.

The dense distribution of retina fields very close to the center of the input image, as seen in Fig. 5, also highlights the negative effect of the underdeveloped vision to the sensorimotor control to explore a broader view. The concentration around the fovea during the first months of development implies that

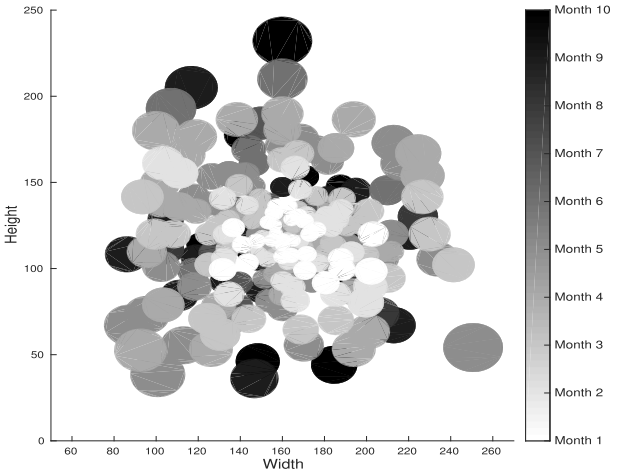

Fig. 5. Retina map population during development. The fields are plotted in reverse to clearly depict the concentration of those generated during early months close to the foveal area. This is an emergent effect of the narrow FOV during early months.

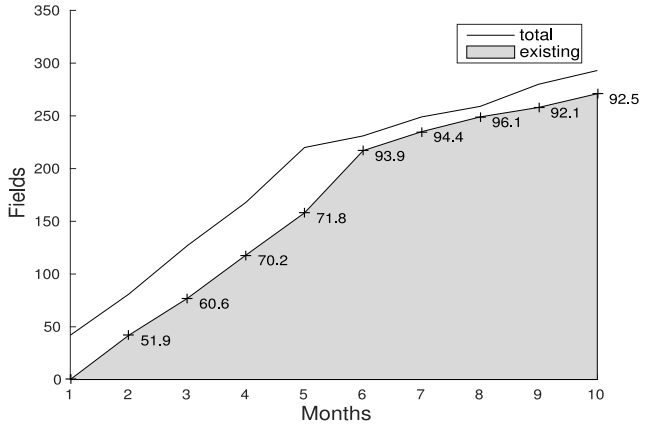

Fig. 6. Usability of existing fields as compared to the total fields created during the ten stages of development in the retina map. The percentage of existing fields used to saccade during each month is also shown.

less eye to motor links are created and confirmed, a result which is also demonstrated in Fig. 4.

The learning progress is also observed when considering Fig. 6. The percentage of retina fields that are reactivated is shown compared to the number of total fields present for each month. The generation of new retina fields is found to reduce in rate after month five. This is explained due to the head which becomes more active as its motor map is gradually populated. Despite the ability to develop the eye control further, the system starts to focus more on acquiring combined eye and head movement experiences. In connection to a wider FOV, saccades are performed by making use of past experiences more often when controlling the eyes, whilst the head contribution is added to the gaze shift. Together, eyes and head allow fixations on targets at the peripheral region. This emergent behavior confirms that defining a saturation type and threshold for the eye control after month three, becomes irrelevant.

The distribution of the retina fields as a result of the narrowed FOV during early months is also seen when considering Fig. 7. As the neck is still locked (months 1–3) and the vision is still underdeveloped (months 4 and 5), retina fields follow an

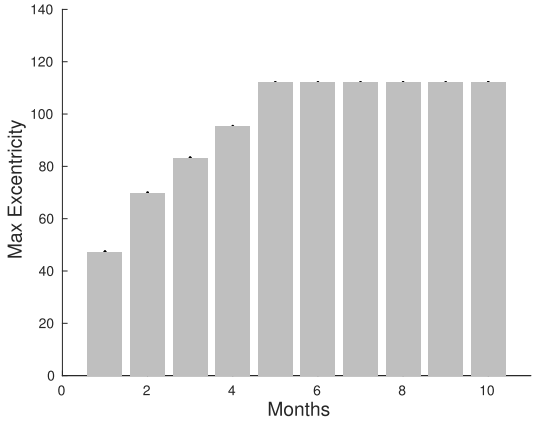

Fig. 7. Maximum distance of retina fields from the center of the input image as a function of the different months of development.

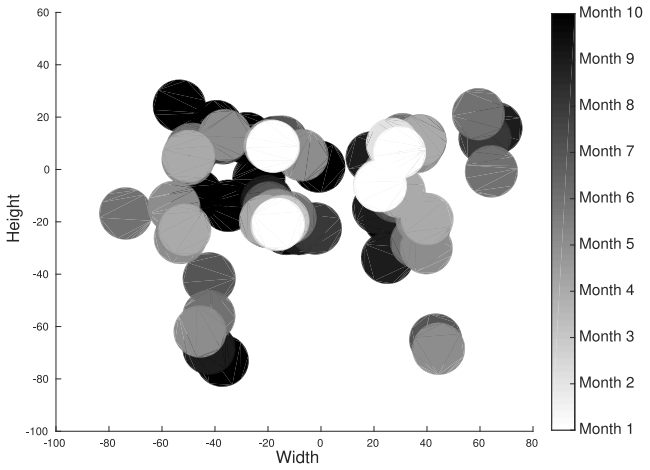

Fig. 8. Gaze space population during early months of development. The system is initially capable of observing features related to regions within its direct environment, as a result of: 1) the immature vision with fewer sensorimotor control links being learned and 2) the head constraint. However, the gaze space is expanded as the system further develops and constraints are lifted.

expanding trend when populating the retina map. After month four, the head starts to contribute thus expanding the retina map more is not necessary.

C. Results on Region Perception

The system's ability to identify stimulating regions in front of the robot and to excite appropriate gaze fields is shown in Fig. 8, where the gaze space is illustrated. The robot is found to focus more on its proximate environment during the first months and gradually expands its understanding about the world as it further develops. The results show that in the first months, the excited fields are closer to the straight ahead direction than in the last months. The average distance from origin is smaller in months 1–3 than 8–10 (1:25.1; 2:24.6; 3:22.5 px; versus 8:36.1; 9:39.3; 10:37.2 px). Again, the underdeveloped vision in terms of FOV and the neck constraints are responsible for limiting the system's ability to explore the surroundings and locate stimulating regions.

Two interesting observations are made when considering the results in Table IV and Fig. 9, regarding the effectiveness of

TABLE IV
LINKS BETWEEN FEATURE TYPES AND GAZE SPACE

Month	1	2	3	4	5	6	7	8	9	10
brightness	18	17	17	10	13	15	11	8	8	21
motion	4	6	1	1	4	2	6	3	2	1
colour	0	0	0	9	12	5	8	3	9	8
edges	0	0	0	1	1	3	2	1	5	6

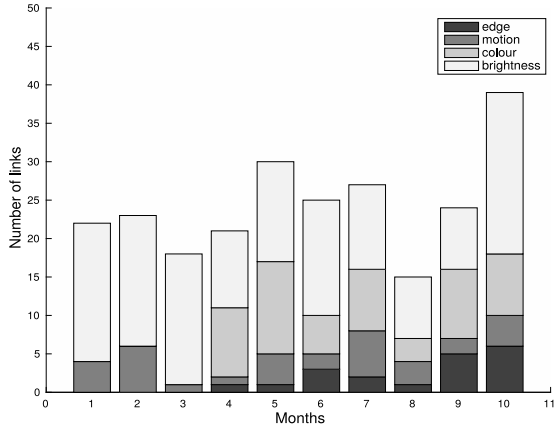

Fig. 9. Feature distribution as a function of developmental stages, using the number of links associated with each feature. The bars depict the total number of links created for each feature of all stimulated regions in the environment.

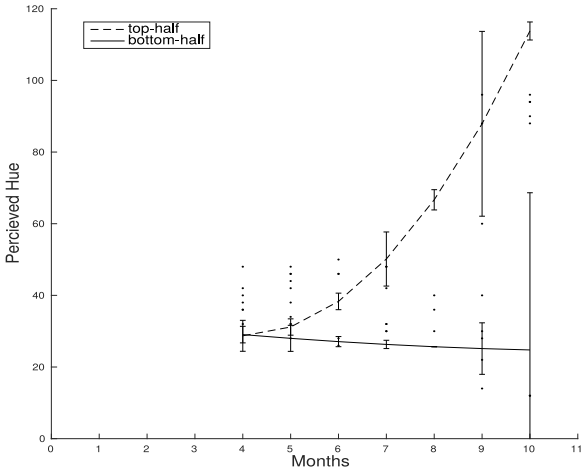

Fig. 10. Mean hue value as a function of months shows the color development. The second degree polynomial fits for the top-half and the bottom-half of the dataset are shown.

feature collection. First, brightness is found to be the most dominant feature, because associated targets can be met on both colored regions and bright regions (e.g., computer monitor). Second, it is observed that despite the existence of colored objects in the environment from the first month of development, color cannot be perceived. This is an emergent behavior of the system and an immediate result of the poor acuity during early stages, which matches findings reported in [15], [29], and [30].

The data collected is analyzed further in terms of color perception. It is observed that color is perceived according to three distinct stages during development. As previously stated, no color perception is achieved during months 1–3, resulting from

TABLE V
EXAMPLES OF TYPICAL FEATURE COLLECTION AS PERCEIVED DURING DEVELOPMENT

			Months									
			1	2	3	4	5	6	7	8	9	10
region 1	Brightness	Value	169.43	166	160.75	153.5	145	128.5	119.25	–	126	197
		Size	801.07	861.17	890.25	984.33	979.5	1011	1096.2	–	1165	1159.5
	Colour	Hue	–	–	–	19.5	21	–	15.333	–	48	–
		Saturation	–	–	–	167.5	170	–	189.67	–	154	–
	Motion	Size	443.25	31.5	73	77	607.5	534	112.5	–	–	–
	Edges	Perimeter	–	–	–	119.6	112.28	114.38	119.36	–	98.698	124.0811

			Months									
			1	2	3	4	5	6	7	8	9	10
region 2	Brightness	Value	183	181.5	175.67	174	168	157.86	146	139	171.67	154
		Size	1352.5	1345.4	1770.2	1611	1834.1	1769.7	2125.5	2429	412.83	688.75
	Colour	Hue	–	–	–	19.2	15.333	17.333	21	19	18.857	30.8
		Saturation	–	–	–	165.6	178	141	121	137	161.71	174.4
	Motion	Size	–	–	–	–	–	211	–	–	–	–
	Edges	Perimeter	–	–	–	–	–	–	–	–	52.385	54.627

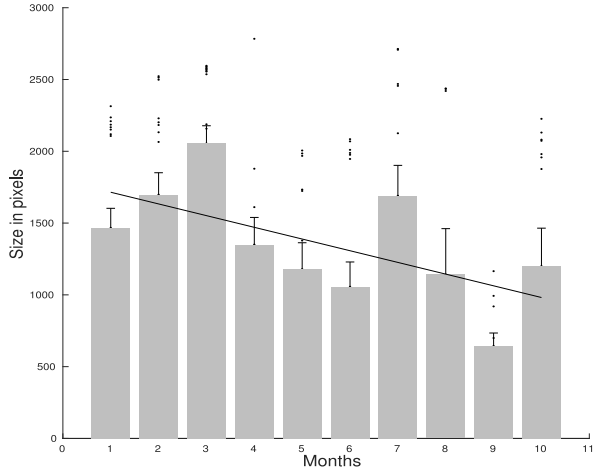

Fig. 11. Size of bright regions as they are perceived during all months of development. The linear fit of the mean values shows that in later months the system is able to detect and observe smaller regions, as compared to early months.

the poor acuity. During months 4–8, color information starts to appear in the system; however, it cannot be used to clearly distinguish between the two main colors presented. The typical range of hue values for red (0–15) and green (90–140) cannot be clearly differentiated, as is shown in Fig. 10. The last stage is seen when considering months 9 and 10, with the distribution of hue values starting to express a bimodal component, gradually increasing the separation between the two ranges of values connected to the gaze fields.

Being a dominant feature, brightness is further analyzed in Fig. 11, where the size of perceived bright regions across all months is depicted. Development is found to have a progressively increasing effect on the ability of the system to accurately understand about sizes of bright regions. The linear fit shows that the system fixates and extracts brightness information on gradually smaller regions. Nevertheless, this result does not necessarily mean that the system perceives new regions each time, rather it is an indication that the same gaze fields are linked to brightness fields which represent more refined data.

Examples of feature data collected when observing two specific regions in the scene across the development of the system are given in Table V. Two gaze fields are selected from the gaze space to reflect two interesting regions in the scene. The fields' collections of features are retrieved using the connecting links that were built at different months of development. Each row summarizes the average values for separate feature quantities. Thus, the table is a low-level illustration of how the system progressively perceived a specific region, based on the features it was able to extract. In the context of comparing and identifying two similar regions in the scene, such data is used to access the two regions' average feature fields in the associated feature maps (e.g., color map) and subsequently to measure the distance between them, along with any overlaps in the feature space.

One aspect to notice is that not all of the features are always present when the region was revisited by the robot. Some of the features are found to remain more constant that others. For example, the edge component (perimeter) seems less variable than color values, which in turn are more constant than brightness and motion values.

This may happen due to: 1) the underdeveloped vision (e.g., the inability to detect colors during first months); 2) the inconsistency of features due to environmental noise; or 3) the region not being revisited as other targets drove the gaze to other directions (as in the case of region 1 in month 8).

Nevertheless, there exists information that stays persistent during all months of development. For instance, when perceived, the perimeter of targets defined by their edges is clearly a strong aspect to consider when describing observed regions. Other similar information is found to be the value of the brightness, which coupled with an also persistent hue and saturation shows a potential for recognizing similarities between features in different regions.

D. Results on Region Recognition

Understanding previously experienced regions at a sufficient level in order to be able to identify similarities between them is an important aspect of the proposed system. As previously

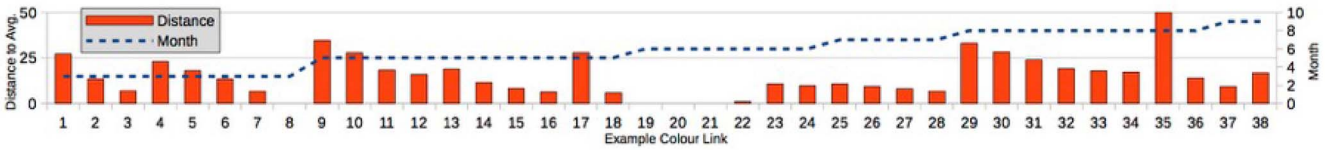

Fig. 12. Variations of acquired color information for a sample gaze field (stimulated region). All links to color information of this particular region are plotted between months 3–9. The height of a bar is the relative distance of newly received information to the current accumulated average in the color space. The dotted line indicates the simulated month in which links are created. Note the peaks at the beginning of a new stage and the decrease within a stage, revealing behaviors related to region recognition.

discussed, the system is further designed to compare known regions and drive its attention toward them while it explores the environment (please see Section III-D). This implies that when the system visually revisits an interesting region, the new feature information that is received is being compared with what the system already has experienced at this region before. Coupled with the memory decay mechanism, by which the system gradually forgets what it knows about a particular region, two behaviors can emerge. First, the system familiarizes itself with a region of interest, if no noticeable changes are depicted. Second, the system can suddenly be aware of rapid changes to a region if the new information is noticeably different from what it already has in its memory.

These two behaviors are both results of recognizing regions, after conducting an experiment with memory decay and comparing similar regions features enabled (Algorithms 2–4). The results are depicted in Fig. 12, where perceiving the color information differences of a particular (sample) region of interest in the scene is shown. The distance of new color information from the average color information built cumulatively in memory captures the familiarity of the system with the region. An increase in this distance implies that the very same region (gaze field) is perceived differently from the weighted average of previously perceived features at this very position. Hence, "unexpected" events are illustrated by the peaks scored at the beginning of each new stage. In more detail, considering the gradual vision development as it is illustrated in Fig. 2, extracted features that are associated with a particular area are perceived differently at each simulated month. In technical terms, when the eyes observe the area of a colored object, the color information that they receive is different from the previous month's observations, leading to the activation of new color fields and their correlation to the area. The new links contribute to the weighted average of color information, therefore the perception of the color is unexpected (peaks), but becomes gradually refined. In fact, the robot can also be "surprised" when the information related to previously observed areas is suddenly perceived differently in the same simulated month, as one calculation of the distance between new and accumulated color information is enough to highlight a substantial change in color. Undoubtedly, observations made in the same month are expected to activate and confirm existing fields and links, respectively, minimizing the divergence in the color perception per observation (decrease of the distance to the average). Indeed, the distance drops as links are created or reused allowing the system to converge to a particular space in the color feature map, as a result of familiarization to the color information it perceives.

V. Conclusion

A major challenge in robotics is the ability to construct stable representations of the world's content and underlying properties from a continuous stream of information. The major challenge of an autonomous system is to be able to learn about region properties with neither explicit supervision, *a priori* knowledge, tuning nor extensive training. In this context, learning ought to be cumulative and incremental, and being triggered only by a primitive set of examples, which are acquired by the system itself.

The system presented in this paper takes inspiration from early infant development where such conditions apply. It extends previous investigations by combining components for eye and head coordination as well as gaze control with feature extraction mechanisms, within an extended design that promotes region perception that aims toward object recognition and use. The use of the mapping system entitles the investigation of the invariance in continuous visual input, and the resulting transparent structures provide the means toward an effective object perception mechanism for embodied systems.

The experiments presented in this paper provide evidence to how the changes in visual inputs can affect the learning of a sensorimotor controller. First, as the vision become more accurate and the FOV widens, the eccentricity of the retina positions linked to a relative eye position increases. The ability to move toward one particular point in space and observe its content progresses rapidly in parallel with a decrease in the amount of steps to perform a saccade between two separated points. Second, the sensorimotor development affects the ability of the system to acquire knowledge about regions from more distant locations. As the vision and the motor control progress, the system becomes able to locate and fixate further peripheral targets and simultaneously collect information of distant targets and their surroundings. Thus, the egocentric gaze space is enriched both in quantity and content (i.e., the correlated feature map values of a particular part of the space). Third, when observing the information describing a region along the development, one has a direct access to its internal projections. Then, in turn, can observe its proximity to previous observations or its proximity to other observations from a different point in space.

Also, the results of an experiment designed to highlight the ability of the system to recognize regions, be familiar and/or surprised with it, depending on the current perception capability is shown. It is shown that with the proposed architecture, the artificial infant is capable of detecting rapid changes at a region, which currently occur due to its developmental changes

in vision, and gradually refine its understanding about the same region while revisiting it.

Another interesting point in these results is the fact that it enables the possibility to drive the future region-related cognition of the robot. For example, if the robot has to perform a visual search of a particular target, it can internally set its own range of tolerance based on the ranges and the consistency of the values it has observed on the previous occurrence of the target and the correlated features that it was able to acquire. Moreover, the system might be able to deliberate on a particular target to decide if it has acquire enough data for being able to identify it later.

The experiments also show that very fast online learning with successful space representation and regional feature acquisition can be performed in less than 5 h. Indeed, longer experiments are now needed to be able to test new hypothesises about how knowledge related to regions of interest can be acquired in real time.

The experiments give a full demonstration of a longitudinal development of both sensorimotor development and early region perception on the iCub humanoid robot. They show how vision development (such as acuity or FOV) can constrain and structure the content of visual information. But more work is now needed to be able to use these results to perform visual search of a previously seen interesting region.

Finally, similarities in color and brightness perception as compared to human findings in literature confirm and validate the choices made in terms of the developmental time-line. This, coupled with the ability of the system to gradually mature and scaffold sensorimotor as well as perception knowledge, enhances the hypothesis that developing a cognitive agent in a stage by stage approach is capable of producing fully autonomous behaviors of recognition as the first step toward gaining affordances and achieving object perception.

Our results show that the robot is able to concurrently discover its sensorimotor abilities and to familiarize itself with the scene, making use of all identified visual features. The memory decay mechanism within months also allows the system to better refine its understanding and, coupled with the similarity mechanism, it becomes aware of visual changes that affect its perception and understanding of the world. When comparing to infant studies, the model and results presented here help to develop an understanding of how infants can start to build up representations of proto-objects based on recognition of similar features in the environment.

A. Future Work

The presented system consists of mechanisms that deal with the simultaneous learning of sensorimotor control and region perception. It is evident that the two functionalities interfere with each other, as the efficiency as well as the internal mechanisms of one affects the other. It is seen that while the increasing development of vision is an important factor for performing successful saccades, the latter is the only available mechanism to drive the attention of the system toward stimulating regions.

Although the simultaneousness of the two functionalities holds for infants during their first months of development, simulating a month in infancy to investigate the details of sensorimotor and cognitive developmental progress is not a straight forward process. Still, both functionalities need to coexist in the same lifespan, a fact that highlights the importance of a design that orchestrates both according to some infant play strategies that will allow the system to spend time and observe interesting regions in the scene. Instead of depending on the learning of sensorimotor control to shift the system's gaze toward some visual stimuli, an excitation approach needs to be employed by which the system will be able to achieve fixations that target to specific regions.

To achieve such an excitation mechanism, several design aspects need to be taken into consideration. Although access to low-level features is given at the current state, the amount of information being collected is not noise-free and thus can easily be misleading. Furthermore, data representing a region is collected through multiple fixations that were not all achieved by a similar head configuration (i.e., eyes' and head's pan and tilt) nor were they subject to the same environmental conditions. Undoubtedly, the way regions are perceived changes, bearing in mind that in a natural environment lighting conditions alter during the day. Thus, mechanisms to organize and validate sets of co-existing and persistent features related to particular regions are important. Once such abstractions and generalized concepts of regions are present, representing proto-objects, they can be used to stimulate saccades and even affect the learning of the sensorimotor control. Furthermore, sequences and coherences identified in the gaze space, attributed to abstractions expressing invariabilities and/or other relative properties (e.g., keeping a similar distance while in motion) are expected to gradually allow the system to achieve a better understanding of objects moving in the scene. This, combined with the ability to interact with these areas, e.g., reaching toward and grasping, is expected to improve the understanding of proto-objects.

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their positive constructive comments helping them to improve this paper.

REFERENCES

[1] J. J. Gibson, *The Ecological Approach to Visual Perception*. Boston, MA, USA: Houghton Mifflin, 1979.

[2] P. Rochat, *The Infant's World* (The Developing Child). Cambridge, MA, USA: Harvard Univ. Press, 2009.

[3] D. H. Hubel and T. N. Wiesel, "The period of susceptibility to the physiological effects of unilateral eye closure in kittens," *J. Physiol.*, vol. 206, no. 2, pp. 419–436, 1970.

[4] D. Maurer, T. L. Lewis, H. P. Brent, and A. V. Levin, "Rapid improvement in the acuity of infants after visual input," *Science*, vol. 286, no. 5437, pp. 108–110, 1999.

[5] J. G. Bremner, A. Slater, and G. Butterworth, *Infant Development: Recent Advances*. Hove, U.K.: Psychol. Press, 1997.

[6] D. Maurer and T. L. Lewis, "Visual acuity: The role of visual input in inducing postnatal change," *Clin. Neurosci. Res.*, vol. 1, no. 4, pp. 239–247, 2001.

[7] M. L. Courage and R. J. Adams, "Visual acuity assessment from birth to three years using the acuity card procedure: Cross-sectional and longitudinal samples," *Optometry Vis. Sci.*, vol. 67, no. 9, pp. 713–718, 1990.

[8] M. S. Banks and P. J. Bennett, "Optical and photoreceptor immaturities limit the spatial and chromatic vision of human neonates," *J. Opt. Soc. America A, Opt. Image Sci.*, vol. 5, no. 12, pp. 2059–2079, 1988.
[9] H. R. Wilson, "Development of spatiotemporal mechanisms in infant vision," *Vis. Res.*, vol. 28, no. 5, pp. 611–628, 1988.
[10] D. L. Mayer *et al.*, "Monocular acuity norms for the Teller Acuity Cards between ages one month and four years," *Invest. Ophthalmol. Vis. Sci.*, vol. 36, no. 3, pp. 671–685, 1995.
[11] V. Dobson, A. M. Brown, E. M. Harvey, and D. B. Narter, "Visual field extent in children 3.5-30 months of age tested with a double-arc LED perimeter," *Vis. Res.*, vol. 38, no. 18, pp. 2743–2760, Sep. 1998.
[12] T. L. Lewis and D. Maurer, "The development of the temporal and nasal visual fields during infancy," *Vis. Res.*, vol. 32, no. 5, pp. 903–911, 1992.
[13] T. L. Lewis and D. Maurer, "Multiple sensitive periods in human visual development: Evidence from visually deprived children," *Develop. Psychobiol.*, vol. 46, no. 3, pp. 163–183, 2005.
[14] A. M. Brown, D. T. Lindsey, E. M. McSweeney, and M. M. Walters, "Infant luminance and chromatic contrast sensitivity: Optokinetic nystagmus data on 3-month-olds," *Vis. Res.*, vol. 35, no. 22, pp. 3145–3160, 1995.
[15] A. M. Brown, "Development of visual sensitivity to light and color vision in human infants: A critical review," *Vis. Res.*, vol. 30, no. 8, pp. 1159–1188, 1990.
[16] D. Y. Teller and M. H. Bornstein, "Infant color vision and color perception," in *Handbook of Infant Perception*, vol. 1. Oxford, U.K.: Wiley, 1987, pp. 185–236.
[17] M. S. Banks and E. Shannon, "Spatial and chromatic visual efficiency in human neonates," in *Visual Perception and Cognition in Infancy*. Hillsdale, NJ, USA: Lawrence Erlabaum Assoc., 1993, p. 146.
[18] R. Baillargeon, "Physical reasoning in infancy," in *The Cognitive Neurosciences*. Cambridge, MA, USA: MIT Press, 1995, pp. 181–204.
[19] E. S. Spelke, K. Breinlinger, J. Macomber, and K. Jacobson, "Origins of knowledge," *Psychol. Rev.*, vol. 99, no. 4, pp. 605–632, 1992.
[20] R. Baillargeon and M. Graber, "Where's the rabbit? 5.5-month-old infants' representation of the height of a hidden object," *Cogn. Develop.*, vol. 2, no. 4, pp. 375–392, 1987.
[21] T. Wilcox, "Object individuation: Infants' use of shape, size, pattern, and color," *Cognition*, vol. 72, no. 2, pp. 125–166, 1999.
[22] S. P. Johnson, "How infants learn about the visual world," *Cogn. Sci.*, vol. 34, no. 7, pp. 1158–1184, 2010.
[23] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.
[24] J. Law, P. Shaw, and M. Lee, "A biologically constrained architecture for developmental learning of eye–head gaze control on a humanoid robot," *Auton. Robots*, vol. 35, no. 1, pp. 77–92, 2013.
[25] P. Shaw, J. Law, and M. Lee, "A comparison of learning strategies for biologically constrained development of gaze control on an iCub robot," *Auton. Robots*, vol. 37, no. 1, pp. 97–110, 2014.
[26] B. Johnston, K. Saunders, and J.-A. Little. (2015). *Visual Acuity Measures Printable Examples*. [Online]. Available: http://biomed.science.ulster.ac.uk/vision/-Visual-Acuity-.html
[27] M. H. Lee, Q. Meng, and F. Chao, "Staged competence learning in developmental robotics," *Adapt. Behav.*, vol. 15, no. 3, pp. 241–255, 2007.
[28] P. Shaw *et al.*, "Babybot challenge: Motor skills," in *Proc. Joint IEEE Int. Conf. Develop. Learn. Epigenet. Robot. (ICDL-EpiRob)*, Providence, RI, USA, 2015, pp. 47–54.
[29] A. Franklin, M. Pilling, and I. Davies, "The nature of infant color categorization: Evidence from eye movements on a target detection task," *J. Exper. Child Psychol.*, vol. 91, no. 3, pp. 227–248, 2005.
[30] D. Allen, M. S. Banks, and A. M. Norcia, "Does chromatic sensitivity develop more slowly than luminance sensitivity?" *Vis. Res.*, vol. 33, no. 17, pp. 2553–2562, 1993.

Alexandros Giagkos received the B.Sc. degree in computer science, the M.Sc. degree in Internet and distributed systems, and the Ph.D. degree in biologically inspired networking from Aberystwyth University, Aberystwyth, U.K.

He is a Post-Doctoral Research Associate with the Intelligent Robotics Group, Aberystwyth University. His current research interests include developmental, evolutionary, and swarm robotics.

Daniel Lewkowicz received the B.Sc. degree in neuroscience from Toulouse University, Toulouse, France, in 2010, and the Ph.D. degree in cognitive psychology from Lille University, Lille, France, in 2013.

He is currently a consultant in cognitive ergonomics for a leading aeronautic company in Toulouse, France, having previously worked as a Post-Doctoral Researcher with Aberystwyth University, Aberystwyth, U.K. His current research interests include interdisciplinary work between neuroscience, psychology, and robotics for a better understanding and modeling of human cognition.

Patricia Shaw received the B.Sc. degree in artificial intelligence and the Ph.D. degree in computer science from the University of Durham, Durham, U.K., in 2005 and 2010, respectively.

She was a Post-Doctoral Research Associate with Aberystwyth University, researching developmental robotics as part of the European Framework 7 IM-CLeVeR project, where she is currently a Lecturer with the Intelligent Robotics Group. Her current research interests include biologically and psychologically inspired architectures for developmental learning in robotic systems.

Suresh Kumar received the B.E. degree in electronics engineering from Mehran UET, Jamshoro, Pakistan, and the M.S. degree in control engineering from GCU, Lahore, Pakistan. He is currently pursuing the Ph.D. degree in intelligent robotics with Aberystwyth University, Aberystwyth, U.K.

He was a Lecturer with Sukkur IBA, Sukkur, Pakistan, and he is now on a study leave. His current research interests include developmental and cognitive robotics.

Mark Lee received the B.Sc. and M.Sc. degrees in electrical engineering from the University of Wales, Swansea, U.K., in 1967 and 1969, respectively, and the Ph.D. in psychology from Nottingham University, Nottingham, U.K., in 1980.

He is currently a Professor of Intelligent Systems with the Department of Computer Science, Aberystwyth University, Aberystwyth, U.K. He was a PI on four recent U.K. Engineering and Physical Sciences Research Council and EC funded research projects on robotic sensory-motor learning, adaptation, and development. His current research interests include developmental robotics, particularly in relation to early infant psychology.

Qiang Shen received the Ph.D. degree from Heriot-Watt University, Edinburgh, U.K., and the D.Sc. degree from Aberystwyth University, Aberystwyth, U.K.

He is the Chair of Computer Science and the Director of the Institute of Mathematics, Physics and Computer Science, Aberystwyth University. He has authored two research monographs and over 350 peer-reviewed papers. His current research interests include computational intelligence, reasoning under uncertainty, pattern recognition, data mining, and real-world applications of such techniques for intelligent decision support (e.g., crime detection, consumer profiling, systems monitoring, and medical diagnosis).

Dr. Shen was a recipient of the Outstanding Transactions Paper Award from the IEEE. He is a long-serving Associate Editor of the IEEE TRANSACTIONS ON CYBERNETICS and the IEEE TRANSACTIONS ON FUZZY SYSTEMS, and an Editorial Board Member of several other leading international journals. He is a fellow of the Learned Society of Wales.